

Ammons Scientific

ORIGINALITY | CREATIVITY | UNDERSTANDING

Comprehensive Psychology

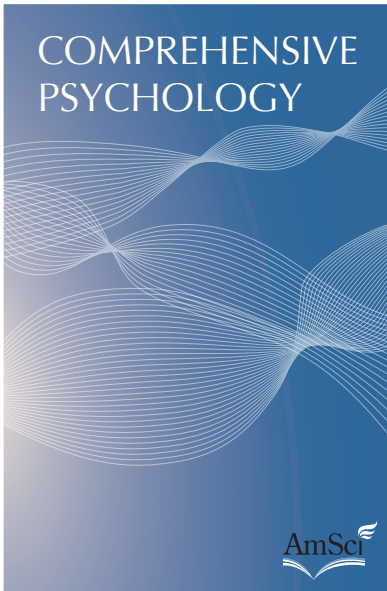
Comprehensive Psychology is an Open Access peer-reviewed publication and operates under the CC-BY-NC-ND Creative Commons License. The Author(s) retains copyright to this article and all accompanying intellectual property rights.

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.'

Additional Information about the terms and conditions of this Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License can be found at www.CreativeCommons.org



CITATION

Kempenaers, C., De Boeck,
P., Dehon, C., Braun, S., &
Linkowski, P. (2014) Metric
invariance of the French
version of the Contingencies
of Self-Worth Scale.
Comprehensive Psychology,
3, 22.

Metric invariance of the French version of the Contingencies of Self-Worth Scale^{1, 2, 3}

C. Kempenaers

Laboratory of Psychiatric Research, Psychiatric Department, C.U.B. Erasmus
Hospital, Université libre de Bruxelles, Belgium

P. De Boeck

The Ohio State University
University of Leuven, Belgium

C. Dehon

ECARES, Université libre de Bruxelles, Belgium

S. Braun, P. Linkowski

Laboratory of Psychiatric Research, Psychiatric Department,
C.U.B. Erasmus Hospital, Université libre de Bruxelles, Belgium

Abstract

The Contingencies of Self-Worth Scale is a self-report questionnaire assessing seven types of 'reward' for self-worth. To test its construct validity, the French version was administered to 787 undergraduate students of both sexes and the data submitted to confirmatory factor analyses (CFA) with the robust MLM or WLSMV estimators. The reliability of the 7-factor structure as well as its sex invariance were checked with the multiple-group factor model for ordered-categorical measures. The correlated 7-factor structure fitted the data in both sex groups separately first, and then in a simultaneous analysis, confirming the adequacy of the measurement model. Metric invariance was also supported. CFA statistical tests, fit statistics, and parameter estimates as well as indices of local ill-fit were analyzed keeping as basic goal to preserve the questionnaire in its original form. Scalar equivalence across gender was not supported. The exploration of method biases, interacting with gender, as well as with alternative models, could allow to test whether different measurement theories might underlie the data gathered with the French version of the CSWS.

Although there is a longstanding and pervasive interest for the self-esteem construct in social and clinical psychology, as well as in health psychology research, the theoretical domain and valid assessment of self-worth remain largely debated and subject to confusion. Hundreds of publications in the scientific literature have accumulated for decades before a regained conceptual and methodological motivation inspired a new research focus on basic aspects of the self-concept (for a review, see Baumeister, Campbell, Krueger, & Vohs, 2003). In recent years, among other important contributions to this field, Crocker and Wolfe (2001) proposed to stress the multi-faceted nature of self-worth and designed a questionnaire to measure the so-called *contingencies of self-worth* (CSWS; Crocker, Luhtanen, Cooper, & Bouvrette, 2003), relying on the assessment of seven domains, or contingencies, hypothesized to be important internal and external sources of self-esteem (see Method for further details). Structural validation of this self-report assessment by way of confirmatory factor analyses along with predictive and both concurrent and discriminant validity testing by Crocker's team make this instrument an interesting tool (Crocker & Luhtanen, 2003; Crocker, *et al.*, 2003; Luhtanen & Crocker, 2005; Park, Crocker, & Kiefer, 2007).

¹Address correspondence to C. Kempenaers, Psychiatric Dept, C.U.B. Erasmus Hospital, 808 route de Lennik, B-1070 Brussels, Belgium or e-mail (chantal.kempenaers@erasme.ulb.ac.be).

²The matrix of variances and covariances is available from the first author on request.

³The authors greatly acknowledge Mrs. Nicole Delvaux, head of the Department of Psychology, for her helpful support and collaboration, as well as the Erasmus Hospital, Université libre de Bruxelles. Guy Davin, Carol Delanaye, and Fabrice Jurysta, at the Laboratory of Psychiatric Research, collaborated for the data collection; their help was greatly appreciated. Yves Rosseel, at the University of Ghent (Belgium), was also involved at an early stage of this study; we thank him for his friendly contribution.

The present study aims to explore the factorial validity of the CSWS translated into French and presented to a group of freshmen students enrolled in various faculties of Belgian French-speaking universities and schools for higher education. At this stage, however, the study does not encompass a direct cross-cultural validation of the questionnaire. Such a direct comparison would imply that data from U.S. samples and from the French-speaking samples are jointly analysed. Instead, an indirect approach was followed, focusing on the current French sample, to investigate whether the same structure is found as in Crocker and Wolfe (2001). An additional interest was possible sex differences in self-esteem. The valid use of a translated questionnaire is conditional on its psychometrical properties, as assessed in the new language/culture. These properties cannot simply be supposed to carry over from the original version in English. Instead, some basic properties, like the internal consistencies of subscales, must be checked but also the factorial structure, in order to conduct valid research in men and women alike. The sex invariance of this factorial model is a prerequisite (Byrne, 2008). In particular, different levels of invariance across sexes must be confirmed for specific inferential statistics on group attributes to be reliable. The plan is therefore to also assess the equivalence of the original factorial model across sex. Checking the extent to which the psychometric properties of the observed indicators are generalizable across groups or over time is the focus of measurement invariance testing. As such, measurement invariance is a necessary precursor to any group comparison regarding the construct in consideration because it assumes that psychological measurement is not affected by a systematic bias, due to sex, for example.

Depending on the results of the measurement invariance across sex, it will also be possible to compare the self-esteem of men and women in the seven domains of the Crocker and Wolfe (2001) questionnaire. Gentile, Grabe, Dolan-Pascoe, and Wells (2009) have published a meta-analysis on domain-specific sex differences regarding self-esteem, but these results may not be automatically be generalized to self-esteem contingencies. Crocker, et al. (2003) found that compared with men, women's self-worth depends more on approval, appearance, academics, and family support, which fits with a rather social profile, in line with findings by Josephs, Markus, and Tafarodi (1992). Similar results were found by Stefanone, Lackaff, and Rosen (2011) and Maricutoiu, Macsinga, Rusu, Virga, and Sava (2012), both with student samples and 10 years later, in the U.S. and Romania, respectively. In the Romanian study, the scores of women were higher overall. Somewhat surprisingly, there was no clear evidence for men's self-worth being more contingent on competition, as could be expected on the basis of Josephs, et al. (1992). Note that, in all these studies, the comparison relies on subscale summary scores,

which is not an optimal method without a study of measurement invariance.

Central to this research as a statistical method is confirmatory factor analysis (CFA). CFA is a type of structural equation modeling (SEM) that deals specifically with measurement models, specifying the relationships between observed measures and latent variables (Bollen, 2002; Brown, 2006). In particular, and in contrast to exploratory factor analysis (EFA), CFA allows testing how well raw data fit an underlying *a priori* model; it is therefore considered particularly appropriate for assessing replication of a given factorial model across different samples and languages, which is part of the validation of any test instrument. Indeed, one of the major specific advantages of CFA is its capability to examine the equivalence of (all measurement and structural) parameters of the factor model across multiple groups.

The particular statistical challenge offered by psychological questionnaire databases resides in the categorical nature of the response variables (Likert-type format). Beauducel and Herzberg (2006) have shown that for the case of a normal underlying variable distribution, one can treat the item scale as continuous from four to five categories on without too much bias. It is therefore common practice in SEM analyses to use either the Maximum Likelihood (ML) estimator or its robust counterpart MLM in case of non-normal data (Byrne, 2012). Working with item parcels would be another solution, but is far from ideal with the inventory under investigation which has a low item-to-factor ratio (see Method). In a multiple group context, as is the case for invariance testing, WLSMV (robust weighted least squares, Muthén & Muthén, 1998–2004) is more specifically needed for CFA with categorical variables. Indeed, as Lubke and Muthén (2004) have demonstrated, an analysis of Likert data in a multiple group context, under the assumption of multivariate normality, may distort the factor structure differently across groups, rendering investigations of measurement invariance problematic. Millsap and Yun-Tein (2004) hence recommended a multiple population extension of the ordered-categorical factor model (with WLSMV) to study invariance across groups (as is, among others, the case for the comparison of a factor structure across sexes). Following these authors' recommendation, it was therefore decided to proceed with this specific adjusted methodology, the multiple group confirmatory factor model for ordered-categorical outcomes, for the invariance testing. However, because most work published in the psychometric field still relies on the continuous methodology (with ML or MLM estimators), and because in-depth discussions of the innovative multiple group mean and covariance structure analysis for ordinal variables remain scarce in the literature, the fit of the measurement model assuming continuous data also will be tested, but with the robust MLM estimator in case of non-normal distributions.

Method

The study was approved by the Ethics Committee of the Medical Faculty of the university the first author is affiliated with.

Participants

Undergraduate students ($n=787$) affiliated with several schools of higher education or university faculties (nursing school, law, economics, psychology, medicine, and engineering), participated on a voluntary basis. Of the 787 participants, 464 (59%) were women. The mean age was 19.3yr. ($SD=1.5$, range=17 to 25). Race/ethnicity was not recorded; participants were all French-speaking. The participants responded to five self-report scales; they recorded their responses on scannable answer sheets with a unique yet anonymous identification code. The assessment session was collective, lasted about one hour, and took place during a scheduled class time.

Measures

The scale was administered in a semi-random order as part of a package of five self-report scales tapping emotional and social attitudes and traits (alexithymia, empathy, autistic functioning, and resilience). The data from these other questionnaires is not analyzed here. A questionnaire assessing some basic demographic features was also presented.

The Contingency of Self-Worth Scale (CSWS, Crocker, et al., 2003) is composed of 35 items belonging to seven subscales (five items per subscale), each tapping one of the following hypothesized domains of self-worth contingencies: Family Support (Factor 1; e.g.: "When my family members are proud of me, my sense of self-worth increases"), Competition (F2; e.g.: "Doing better than others gives me a sense of self-respect"), Appearance (F3; e.g.: "When I think I look attractive, I feel good about myself"), God's Love (F4; e.g.: "My self-worth is based on God's love"), Academic Competence (F5; e.g.: "My self-esteem is influenced by my academic performances"), Virtue (F6; e.g.: "My self-esteem would suffer if I did something unethical") and Others' Approval (F7; e.g.: "My self-esteem depends on the opinions others hold on me"). Five contingencies (F1, F2, F3, F5, and F7) are considered as measuring "external" sources of self-worth and the two others (F4 and F6) as measuring "internal" sources. A Likert-type scale with seven anchor points is used for responding, with anchors 1: Strongly disagree and 7: Strongly agree. Half the items are reverse coded. Each subscale score (the sum of its five item scores) provides a measure of the corresponding contingency of self-worth. In its original English, the scale showed good psychometric properties, with evidence supporting its factorial structure through CFA, high internal consistencies for the seven subscales, and interesting results concerning convergent and

discriminant validity as well as sex differences (Crocker, et al., 2003).

The translation of the scale was a three-step process, following the usual procedure for the translation of questionnaires: (1) the items were translated into French by the first author with an emphasis on conceptual and cultural rather than linguistic equivalence; (2) a native English-speaking psychologist with fluency in French then translated all items back into English; and (3) the first author and the English native speaker finally compared the original and back-translated versions of the questionnaire and, when needed, adapted the French items to fit the precise meaning of the original English version better. See Table 1 for the final French version of the questionnaire.

Analyses

Mplus version 5.21 (Muthén & Muthén, 1998–2009) was used for CFA to assess the factorial validity of the proposed seven-factor structure of the scale as well as the measurement invariance. SPSS Versions 17 through 19 (2008–2010) were used for other statistical analyses.

In order to compare data to Crocker's, subscale means (SD) were computed for men and women apart. Cronbach's α was also computed for the 7 subscales (men and women together).

Confirmatory Factor Analysis.—The purpose of CFA, like EFA, is to identify latent factors that account for the variation and covariation among observed measures. They are both based on the *common factor model* which postulates that each indicator in a set of observed measures is a linear function of one or more common factors and one unique factor. Although EFA and CFA both aim to reproduce the observed relationships among indicators with a smaller set of latent variables, they differ by the number and nature of *a priori* specifications and restrictions made in the factor model. In CFA, a pre-specified factor solution – in term of number of factors, pattern of indicator-factor loadings, independence or covariance of the factors and indicator unique variances – is evaluated in terms of how well it reproduces the sample correlation or covariance matrix of the measured variables (Brown, 2006). The objective of CFA is therefore to obtain estimates for each parameter of the measurement model allowing the predicted variance-covariance matrix to reproduce the sample variance-covariance matrix as well as possible.

One of the most commonly used statistical methods for estimating parameters of the common factor model is the Maximum Likelihood procedure (ML), which relies on the assumption of multivariate normality. With non-normal data, and especially for kurtotic variables, mean-adjusted maximum likelihood (MLM) is preferred as a maximum likelihood estimator with robust standard errors and provides a corrected (scaled) χ^2 statistic, better known as Satorra-Bentler χ^2 (Byrne, 2012).

TABLE 1
Content of the French Version of the CSWS (by Subscale and with Item Number in First Column)

Family Support	
7	Savoir que les membres de ma famille m'aiment me renvoie une impression positive de moi-même
10	Ma valeur personnelle n'est pas influencée par la qualité de mes relations avec les membres de ma famille
16	Quand les membres de ma famille sont fiers de moi, la perception que j'ai de ma valeur personnelle augmente
24	Quand je ne me sens pas aimé(e) par ma famille, mon estime de moi diminue
29	Il est important pour mon propre respect d'avoir une famille qui se soucie de moi
Competition	
3	Je sens que j'ai de la valeur lorsque je réussis une tâche mieux que les autres personnes
12	Savoir que je réussis une tâche mieux que les autres augmente mon estime de moi-même
20	Faire mieux que les autres me procure un sentiment de respect personnel
25	Ma valeur personnelle est affectée par mon niveau de réussite quand je suis en compétition avec d'autres
32	La perception de ma valeur personnelle est influencée par la manière dont je réussis des tâches qui impliquent une compétition avec d'autres personnes
Appearance	
1	Quand je pense être séduisant(e), je suis content(e) de moi
4	Mon estime de moi-même n'est pas liée à l'idée que j'ai de mon apparence extérieure
17	Mon estime de moi est influencée par la manière dont je trouve mon visage ou mes traits attrayants
21	La sensation que j'ai de ma valeur personnelle en souffre quand je pense que je n'ai pas l'air bien
30	Mon estime personnelle ne dépend pas du fait que je me trouve séduisant(e) ou pas
God's Love	
2	Ma valeur personnelle est basée sur l'amour de Dieu
8	Je sens que j'ai de la valeur quand j'ai l'amour de Dieu
18	Si je n'avais pas l'amour de Dieu, mon estime de moi-même en souffrirait
26	Mon estime de moi-même augmente quand je sens que Dieu m'aime
31	Quand je pense que je désobéis à Dieu, je ne suis pas content(e) de moi
Academic Competence	
13	Mon opinion de moi-même n'est pas liée à la manière dont je réussis à l'école/université
19	Bien travailler à l'école/université me procure un sentiment de respect envers moi-même
22	Je suis davantage content(e) de moi quand je sais que j'ai bien travaillé à l'école/université
27	Mon estime de moi-même est influencée par mes performances académiques
33	Je ne suis pas content(e) de moi quand mes résultats académiques/scolaires ne sont pas satisfaisants
Virtue	
5	Faire quelque chose que je trouve mal me fait perdre le respect de moi-même
11	Quand je suis mes principes moraux, le respect que j'ai pour moi-même est renforcé
14	Je ne pourrais pas me respecter si je ne vivais pas selon un code moral établi
28	Mon estime de moi-même en souffrirait si je faisais quelque chose qui n'était pas éthique/moral
34	Mon estime de moi-même dépend du fait que je suis ou non mes principes éthiques/moraux
Others' Approval	
6	Je m'en fiche si les autres ont une opinion négative de moi
9	Je ne peux pas me respecter si les autres ne me respectent pas
15	Je me fiche de ce que les autres pensent de moi
23	Ce que les autres pensent de moi n'a pas d'influence sur ce que je pense de moi
35	Mon estime de moi-même dépend de l'opinion que les autres ont de moi

Mean- and variance-adjusted least squares (WLSMV), is the best option for models for ordered-categorical outcomes or severely non-normal data (Brown, 2006). It provides weighted least squares parameter estimates using a diagonal weight matrix and robust standard errors and a mean- and covariance-adjusted χ^2 test statistic. The MLM was used, as is most common, for all tests of model fit; WLSMV was preferred for the invariance testing to minimize errors due to non-normal data in a multiple group analysis, such as with sex groups. Raw data were used as input for the analyses. Scaling of the latent variables was set by fixing the loading of the first item of each subscale to 1.

The hypothesized original measurement model with 7 factors assumed the following. (1) Responses are explained by seven factors or contingencies of self-worth: Family Support, Competition, Appearance, God's Love, Academic Competence, Virtue, and Others' Approval. (2) Each subscale item has a nonzero loading on the factor that it is supposed to measure and a zero loading on the other six factors. (3) The seven factors were allowed to correlate. (4) Error variances were uncorrelated.

The 7-factor model was tested against alternative models, as in the Crocker study (Crocker, et al., 2003): a 1-factor model (all items measuring a global self-worth dimension), a 2-factor model (internal vs. external sources of self-worth), and a 3-factor model with three factors based on the wording of the items ("self-esteem goes up if...", "self-esteem goes down if...", or self-esteem depends on..."). The total student group was randomly split in two halves to run the comparative CFA analyses of the four competing models in the two halves of the sample. The best model of the four was chosen, in terms of goodness of fit and consistency across sample halves, to continue with a CFA for categorical variables (WLSMV) and to investigate the measurement invariance across the sex groups with a multigroup CFA.

Chi-squared and fit indices.—The basic test of goodness-of-fit is the chi-squared test, which is used for evaluating the fit of a pre-specified model to the actual data covariance/correlation matrix. Because of its inherent limitations, the chi-squared test statistic is usually supplemented by a series of additional indices (Brown, 2006), among which the following are the most commonly used ones (Jackson, Gillaspay, & Purc-Stephenson, 2009). The first index is the Comparative Fit Index (CFI; Bentler, 1990), for which a value of 0.90 or higher indicates a reasonable model fit and a value of 0.95 indicates a good fit. The second index is the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), for which values approaching 1.0 (greater than 0.95) are interpreted as indicating a good fit. The Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1993) for which a value of 0.06 or lower indicates a good fit is a third widely used index of fit. The CFI and TLI are comparative (relative) indices, comparing an estimated structure with a random

structure, and the RMSEA is an index of absolute goodness of fit to the data. Other guides to model evaluation (extent and sources of [mis]fit) rely on the examination of model estimated parameters and inspection of normalized residuals and modification indices provided by *Mplus*. Despite the acknowledged interest of rules of thumb regarding the interpretation of fit indices, several researchers increasingly warn against their poor generalizability to any given research data set, and this issue is hotly debated (e.g., Marsh, Hau, & Wen, 2004; see Special Issue about SEM in *Personality and Individual Differences*, 2007, Volume 5). The interpretation of results will therefore rely on a combination of multiple information sources with a substantively- and methodologically-based analysis.

Invariance testing with CFA.—A test or questionnaire is measurement invariant if the distribution of observed scores conditional on the factor scores is the same for all groups, which means that the regression relations between observed items and underlying factors have to be the same across groups (Meredith, 1993). In the case of continuous data, the measurement parameters under scrutiny are the intercepts, factors loadings, and residual variances, but the set is different when ordered-categorical data are concerned (Lubke & Muthén, 2004) and includes so-called "thresholds." These thresholds are model parameters that delineate the response categories, and they can differ depending on the item. For more details on this topic, which extends beyond the scope of the paper, the reader is referred to Brown (2006) and to the *Mplus* user guide (Muthén & Muthén, 1998-2004) as well as to the Muthén and Asparouhov *Mplus* Web note (2002). Besides, for a detailed presentation of the model specification and identification constraints in this framework, see Millsap and Yun-Tein (2004) and van der Sluis, Vinkhuyzen, Boomsma, and Posthuma (2010).

The successive steps of measurement invariance testing followed are summarized as follows. *Configural invariance* tests whether the (item) content of the 7 *a priori* factors, whatever their respective loadings, is the same for men and women (Model 1 or "baseline model"). Configural invariance does not exclude that what seems to be similar factors needs to be interpreted differently depending on the sex. *Metric invariance* requires that the relations between the items and their latent factor are the same across sex (meaning that their loadings are equal; Model 2). For *strong factorial invariance* to hold, also *scalar invariance* must apply, meaning that the thresholds have to be identical in the two sex groups (Model 3); this allows interpretation of sex differences in factor means. *Strict factorial invariance* means that also the error variances of the observed items (their part unrelated to the latent factor) are equal across groups (Model 4).

The procedure for invariance testing requires that the pre-specified measurement model be first tested

separately in the two groups at hand (men and women), to assess the goodness-of-fit of the model to the actual samples before progressively evaluating the requirements of invariance in the simultaneous sex-based CFA (Models 1 through 4). Aside evaluating the goodness-of-fit of each of the increasingly constrained models with the chi-squared test statistic and fit indices, invariance testing also relies on the difference chi-squared test statistic, or likelihood-ratio test, which assesses whether a *nested* (more restrictive) model differs from a *parent* (less restrictive) model. To obtain the correct chi-squared difference test with the WLSMV estimator, a two-step procedure is required with *Mplus* (DIFFTEST option). Testing for sex invariance hence requires comparing increasingly constrained models in a step-wise manner, from equal loadings in men and women first, to equal loadings and thresholds, and at last equal loadings, thresholds, and unique variances. Note that different factor means and variances depending on the group are not a violation of measurement invariance. Such differences rather reflect genuine group differences and not measurement biases.

The theta parameterization with standardized loadings was applied to study invariance (Millsap & Yun-Tein, 2004). As a consequence, the loadings may seem to differ between groups even when they would in fact be equal if the variances would not have been scaled to 1.00. This option was chosen because this is how factor models are commonly presented.

Results

At an item level, 0.25 to 2.4% responses were missing, corresponding to 2 to 19 of 787 participants not responding to at least one item. Further analyses used all available complete data with listwise deletion.

Distribution of Item Responses

Examination of response distributions as well as skewness and kurtosis statistics showed that items exhibiting a normal-like distribution were the exception. Most items showed a moderate to high endorsement (item

distributions skewed to the left) with a very significant exception for all five items measuring the God's Love factor. Indeed, most participants denied the relevance of God's love to their self-worth (most responses for "Strongly disagree"), while a significant minority gave a neutral answer (item distributions skewed to the right). In conclusion, many items showed a skewed distribution, and for God's Love even a bimodal distribution was observed. Kurtosis also was clearly problematic for a large majority of items.

Internal Consistency and Descriptive Statistics

Except for the Appearance subscale, the internal consistency of the CSWS was good to excellent (see Table 2). However, the internal consistency was somewhat lower for several subscales compared to Crocker's (Crocker & Wolfe, 2001).

The mean subscale scores were lower in the current sample compared with Crocker and Wolfe (2001). It indicates that in general, the current participants' self-worth had a lesser extent of contingency and thus less dependence. This may reflect cultural or time differences in sensitivity of self-worth. Especially for God's Love the means were lower, which is in line with the declining trend of religion in Europe. The means indicate that Academic Competence, Family Support, Virtue, and Competition were the most important sources of self-worth in the current sample.

Evaluation of 7-Factor Model (Robust Continuous Methodology – MLM Estimator)

As shown in Table 3 (Part A), the 7-factor model is clearly the best among the competing models in the two half samples. Only in this model did all three fit indices, CFI, TLI, and RMSEA, attain reasonable values, although the results for the second half show somewhat lower goodness-of-fit. Grouping thereafter the two halves into one total group, about the same results were obtained for the 7-factor model (see part B of the Table). This supports the 7-factor model as clearly better than the other three models. The separate CFA results for men and women (Part B of Table 3) globally indicate

TABLE 2
Means (SD) and Internal Consistencies for the 7 Subscales of the French CSWS

Subscale	Women ^a		Men ^a		Total		Cronbach's α ^b
	M	SD	M	SD	M	SD	
1. Familial Support	5.2	0.05	4.9	0.06	5.1	1.05	75
2. Competition	4.7	0.06	4.8	0.07	4.8	1.25	88
3. Appearance	4.5	0.05	4.2	0.06	4.4	1.03	67
4. God's Love	2.6	0.09	2.4	0.10	2.5	1.81	96
5. Academic Competence	5.4	0.05	5.2	0.07	5.3	1.09	82
6. Virtue	5.0	0.05	4.8	0.06	4.9	1.08	74
7. Others' Approval	4.1	0.06	3.7	0.08	3.9	1.35	79

Note. — ^aSample sizes range from 437 to 443 in women, 303 to 314 in men. ^bSample sizes range from 741 to 753.

TABLE 3
Confirmatory Factor Analysis: Fit Statistics and Indices for
Several Measurement Models for CSWS

(A) Comparison of alternative models in the first ($n = 339$) and second half samples ($n = 334$) ^a				
Models (in first half sample)	^b Robust χ^2	CFI	TLI	RMSEA
7 factors ($df = 539$)	877.95	0.934	0.927	0.043
3 factors ($df = 557$)	3875.79	0.351	0.306	0.133
2 factors ($df = 559$)	2308.12	0.658	0.636	0.096
1 factor ($df = 560$)	4010.92	0.325	0.283	0.135
Models (in second half sample)				
7 factors	1144.96	0.880	0.868	0.058
3 factors	3827.92	0.355	0.310	0.133
2 factors	2528.78	0.611	0.586	0.103
1 factor	3994.44	0.322	0.280	0.136
(B) Baseline 7-factor model in men and women and in total sample				
Group	Robust χ^2	CFI	TLI	RMSEA
Females ($n = 395$)	1165.61	0.892	0.881	0.054
Males ($n = 278$)	918.07	0.908	0.899	0.050
Total ($n = 681$)	1339.20	0.919	0.911	0.047

Note. — ^aDue to robust methodology, sample sizes result from LISTWISE selection. ^bAll χ^2 significant, $p < .00001$. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation.

a good fit of the 7-factor model, although the comparative fit indices (CFI and TLI) had only marginal values. This may be the consequence of not using the categorical approach here. Because the WLSMV approach is recommended (see earlier) for multigroup models with categorical data, and because the 7-factor model seems to be clearly better than the other three, the investigation of measurement invariance was continued with the 7-factor model and a WLSMV approach.

Gender Invariance Testing (Ordered-Categorical Methodology – WLSMV Estimator)

The results for the *configural invariance* test across sex (baseline model) are shown in Table 4. CFI and TLI values were clearly improved (0.95 and 0.97, respectively), but the RMSEA (0.094) was above the critical 0.06 value. Not surprisingly because of the large sample size, the chi-squared

value was significant. By way of exploring the misfit, the modification indices as well as the residuals for covariances and thresholds (differences between the model-based estimates and data-based statistics) were inspected; two sources of misfit emerged. The first was the *a priori* imposed absence of cross-loadings in the measurement model. The modification indices indicate that this is too strict a constraint for the relationship between factor 2 (Competition) and several items from subscale 1 (Family Support) and subscale 3 (Appearance). This finding was confirmed by an exploratory principal component analysis (PCA), where also items from other subscales loaded on the Competition factor (e.g., Item 16: “Members of my family are proud of me”; Item 35: “The opinion others have of me contributes to my self-esteem”). Residual covariance seemed to be the second source of misfit. The majority of items with an excessive residual covariance were those defining subscales 3 (Appearance) and 7 (Others' Approval). The two sources of misfit being related, it did not come as a surprise that subscale 3 was involved again. These results can be interpreted as specific forms of family support and others' support being contingent on competition results, while it makes also sense that specific aspects of appearance are related to aspects of others' approval. Apparently there are item-specific relationships that are not captured by the factor correlations, possibly due to specific cultural differences. For example, American parents of college students may be more uniformly proud of their sons and daughters when they do well in competitive activities, and less difference in this respect may have led to a smaller correlation of the parents' pride item with the Competition factor.

However, when considering individual items, not much convergence existed between the two sources of misfit. Hence, because there was no sufficiently stable basis for a modification of the subscales or for the model to be adjusted, it was concluded that the constraints should not be relaxed by incorporating *ad hoc* cross-loadings or error correlations. Although such adjustments could make sense, the indications were not sufficiently strong and convergent to deviate from the original structure validated in earlier studies.

Model misspecification is especially a problem for the absolute goodness-of-fit (less so for the comparative

TABLE 4
Confirmatory Factor Analysis – Progressive Gender Invariance Testing of the 7-Factor Model for CSWS

Models	χ^2	df	Models	$\Delta\chi^2$	df	CFI	TLI	RMSEA
Model 1: baseline (<i>configural invariance test</i>)	879.55‡	197				0.955	0.975	0.094
Model 2: equal loadings (<i>metric invariance test</i>)	871.06‡	200	1 vs 2	34.84†	18	0.955	0.975	0.093
Model 3: equal loadings + thresholds (<i>scalar invariance test</i>)	877.11‡	221	2 vs 3	138.31‡	74	0.956	0.978	0.087

† $p < .01$.

‡ $p < .001$.

goodness-of-fit) when the sample is large (Marsh, *et al.*, 2004; Chen, Curran, Bollen, Kirby, & Paxton, 2008; Bentler, 2010) and when the percentage of unique variance is rather low (Browne, MacCallum, Kim, Andersen, & Glaser, 2002; Stuiwe, 2007). The current sample was large ($N=787$) and the percentage of unique variance was rather low compared with other structures (a moderate value of 50% corresponding to loadings of .70). In a study of CFA and cross-loadings comparable to these, with 49% of unique variance, Stuiwe (2007) showed that a RMSEA value of about 0.10 could be expected instead of a critical value of 0.06, the latter being too low also according to Prudon (2011). The RMSEA value of 0.094 may therefore be no sufficient reason to reject the somewhat misspecified model as unacceptable. Indeed, Marsh, *et al.* (2004) and Chen, *et al.* (2008) have also argued that the critical values proposed by Hu and Bentler (1999), not originally meant by their authors to represent a gold standard, can definitely not be considered as universal criteria. Moreover, misspecified models are not *per se* a necessary reason for rejection in the domain of SEM (MacCallum, 2003; Marsh, *et al.*, 2004; Bollen, *et al.*, 2007; Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007). Finally, not much is empirically known about the behavior of the RMSEA as an index of goodness-of-fit for categorical data, and the problem of misspecification has not yet been investigated for categorical data in a CFA context (Kaniskan, 2011). Moreover, the clearly higher RMSEA with categorical data might also be the consequence of fewer degrees of freedom in the categorical data analysis. For all these reasons, it was decided to continue with a somewhat misspecified model rather than adapting the model on an *ad hoc* basis, especially because the model exhibited a very good relative goodness of fit.

Testing next the *metric invariance* of the measurement model by constraining loadings to be identical for the two sexes would offer an important check of the line followed hitherto in this analysis. In addition to the difference chi-squared test comparing Model 2 to Model 1, fit indices not deteriorating (CFI not decreasing and RMSEA not increasing) would not only mean that the constraint of equal loadings is acceptable but would also confirm that the lack of convergence between misfit indications in the two sexes as reported earlier was not due to substantial systematic sex differences. As seen in Table 4, the CFI value of the metric invariance model (Model 2) was 0.955 and the RMSEA value 0.093, fulfilling the condition just formulated. Furthermore, the difference chi-squared test comparing this model to the baseline model (Model 2) had a p value of .0099 [$\chi^2(18)=34.84$] which is practically 0.01; this value may be considered as borderline and no sufficient basis to reject metric invariance, certainly not in the light of the relative goodness-of-fit values almost unchanged compared to those of the baseline model. Table 5 shows the

TABLE 5
Completely Standardized Parameter Estimates for the Equal Loading Model

Factor and Items	Women ^a		Men ^a	
	Loading	SE	Loading	SE
Family Support (F1) by				
Item 7	0.62	0.03	0.69	0.03
Item 10	0.41	0.04	0.36	0.05
Item 16	0.88	0.03	0.86	0.03
Item 24	0.69	0.04	0.63	0.04
Item 29	0.64	0.03	0.67	0.04
Competition (F2) by				
Item 3	0.815	0.018	0.833	0.020
Item 12	0.878	0.014	0.920	0.014
Item 20	0.832	0.017	0.861	0.017
Item 25	0.756	0.021	0.730	0.027
Item 32	0.742	0.024	0.763	0.023
Appearance (F3) by				
Item 1	0.380	0.048	0.441	0.051
Item 4	0.463	0.043	0.369	0.048
Item 17	0.739	0.031	0.779	0.031
Item 21	0.621	0.040	0.642	0.041
Item 30	0.685	0.030	0.689	0.037
God's Love (F4) by				
Item 2	0.944	0.007	0.920	0.012
Item 8	0.976	0.004	0.966	0.007
Item 18	0.921	0.009	0.910	0.014
Item 26	0.972	0.004	0.947	0.010
Item 31	0.933	0.009	0.907	0.014
Academic Competence (F5) by				
Item 13	0.591	0.029	0.605	0.035
Item 19	0.727	0.025	0.828	0.021
Item 22	0.794	0.021	0.797	0.026
Item 27	0.843	0.019	0.892	0.019
Item 33	0.728	0.026	0.699	0.033
Virtue (F6) by				
Item 5	0.598	0.038	0.519	0.039
Item 11	0.599	0.039	0.727	0.041
Item 14	0.379	0.045	0.532	0.040
Item 28	0.875	0.029	0.735	0.033
Item 34	0.744	0.031	0.787	0.033
Others' Approval (F7) by				
Item 6	0.815	0.019	0.795	0.027
Item 9	0.347	0.041	0.377	0.049
Item 15	0.904	0.015	0.855	0.025
Item 23	0.724	0.026	0.616	0.038
Item 35	0.755	0.024	0.840	0.028

(continued on next page)

TABLE 5 (Cont'd)
Completely Standardized Parameter Estimates for the Equal Loading Model

Factor	Women	Men	
	M	M	SE
Factor 1	0.00	-0.10	0.89
Factor 2	0.00	0.89*	0.42
Factor 3	0.00	0.43	0.51
Factor 4	0.00	-0.20*	0.08
Factor 5	0.00	-0.40	0.46
Factor 6	0.00	1.22*	0.49
Factor 7	0.00	-0.44	0.30

Note. SE = standard error. *All parameters are significant, $p < .001$, except when otherwise specified. * $p \leq .05$; ns = not significant ($p > .05$).

parameter estimates of Model 2 (equal loading model). Please note that: (a) while not allowed to differ and being similar in the unstandardized solution (not shown), loadings may eventually differ in the standardized solution; (b) for model identification reasons on the other hand, factor means were set at zero in women.

All estimated loadings were not only statistically but also quantitatively significant, their values ranging from .347 to .976. The status of the God's Love subscale was peculiar again, with all five items being highly saturated (all loadings above 0.900) and with the lowest standard errors. Three out of seven contingency domains of self-worth showed a significant between-sex difference in estimated means: men scored higher for Virtue and Competition, and lower for God's Love ($p < .05$ for each of these three factors). These results are different from those obtained earlier with the same questionnaire (as explained in the introduction), and they will be discussed in the final section.

Only four estimated factor covariances out of 42 (Table 6) were not significantly different from zero, crediting the *a priori* inter-correlated seven-factor model. In some instances though, the shared variances was very small or virtually null: the God's Love subscale was a somewhat independent factor in both sexes. For most

between-factor covariances, the correlations were a little stronger in men.

Putting a further constraint on the measurement model by equating not only loadings but also thresholds in the two sex groups to test for *scalar invariance* (Model 3) left the CFI almost unchanged and decreased the RMSEA value somewhat, but produced a significantly deteriorated model compared to Model 2 [$\Delta\chi^2(74) = 138.31$, $p < .001$]: strong factorial invariance could not be assumed. The model with added equal unique variances (*strict invariance*; Model 4) was therefore not tested. To explore one possible source of scalar invariance violation at the level of item wording, the authors explored the categorization of questionnaire items (Crocker, et al., 2003) conceived to balance self-worth contingent on success experiences ("positive contingency items"; e.g., "I feel worthwhile when I perform better than others on a task or skill") with self-worth contingent on failure experiences ("negative contingency items"; e.g., "I can't respect myself if others don't respect me") and unspecified self-worth contingencies (e.g., "My self-esteem is influenced by my academic performance"). Scalar differences between sexes, independent of factor means, showed a systematic basis related to the positive vs. negative nature of the contingencies: women agreed more than men with negative contingency items and less than men with positive contingency items.

Discussion

Though SEM has long been used in the statistical field, its implementation and adequate use in psychometric research has significantly expanded only in the last 10 to 15 years. Although they may not be overemphasized, several methodological topics still remain controversial in CFA, and more specifically the reliance on cut-off values for fit indices in evaluating factorial models. The recommendations now insist on evaluation being based on multiple and complementary (but possibly conflicting) indices as well as on the combination of substantive and statistical evidence (Tomarken & Waller, 2005;

TABLE 6
Estimated Factor Covariances of the Completely Standardized Solution for Equal Loading Model

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Factor 1		0.40‡	0.33‡	0.21‡	0.57	0.39‡	0.32‡
Factor 2	0.44‡		0.52‡	0.11*	0.70	0.22‡	0.26‡
Factor 3	0.51‡	0.57‡		-0.12*	0.43‡	-0.02	0.51‡
Factor 4	0.24‡	0.10	0.12*		0.11*	0.25‡	-0.07
Factor 5	0.58‡	0.62‡	0.40‡	0.15*		0.36‡	0.35‡
Factor 6	0.40‡	0.23‡	0.09	0.29‡	0.42‡		0.04
Factor 7	0.39‡	0.33‡	0.49‡	0.13*	0.32‡	0.20‡	

Note.—Figures for women above the diagonal, for men below.

* $p < .05$.

‡ $p < .001$.

Jackson, et al., 2009). Another basic aspect of CFA methodology that is diversely handled by researchers is the need to adapt a specific analytical framework for ordered-categorical variables, as is the case with questionnaire data and their Likert response format.

In this study, a Belgian French self-report questionnaire measured dimensions of self-esteem. In the early 2000s in the USA, the work of Jennifer Crocker to design a measure of self-worth contingencies (Crocker & Wolfe, 2001) was particularly promising because of the heuristic utility of the self-worth concept in psychology, on the one hand, and the highly confusing psychometric field around the measurement of self-esteem at that time, on the other hand. Moreover, her publications on the original American English scale further credited the questionnaire as a valid tool, regarding both its construct, convergent, and predictive validity in student groups (Crocker & Luhtanen, 2003; Park & Crocker, 2005; Park, et al., 2007). Particularly interesting was the conceptualization of self-worth relying on varied distinct contingencies, from more “internal” ones (God's Love, Virtue) to more “external” ones (Others' Approval, Appearance), through others relating to Family Support, Competition, and Academic Competence.

Translation does not automatically produce a ready-to-use instrument. The factorial structure and measurement invariance need to be investigated for validation purposes (Byrne & Watkins, 2003; Gregorich, 2006). Indeed, critical for research as well as for clinical applications is that the scale measures the same construct in different groups. In particular, scalar invariance, i.e., invariance of the factor loadings and thresholds across groups, is required for valid comparisons of latent variable means. Therefore, the measurement model underlying the translated CSWS was tested and, in particular, its sex invariance with an ordered-categorical adjusted CFA based on a two-fold rationale: (a) the CSWS items are formally categorical variables, and (b) omitting deviating items (bimodal or significantly deviated from normality) was not a reasonable option for the translated scale.

The original 7-factor structure of self-worth, with adequate internally consistent dimensions, was largely superior to three alternative models to fit data in men and women. The 7-factor model could moreover be considered equivalent in both sexes, as suggested by excellent comparative goodness-of-fit values. However, as a rather high RMSEA highlighted, this model was somewhat misspecified, namely in terms of cross-loadings and correlated errors, but these sources of misfit did not provide a sufficient basis for adapting the factor model and for deviating from its well-known structure. The question still is how much of misspecification or inaccuracy would render the model inappropriate. The usual guidelines regarding fit indices are largely disputed and some authors clearly allow figures as high as 0.10 as upper limits for RMSEA (Browne & Cu-

deck, 1993) for data similar to these. As already mentioned, the RMSEA seems to be particularly sensitive to misspecification in cases of large samples (Marsh, et al., 2004; Chen, et al., 2008; Bentler, 2010) and low unique variance (Browne, et al., 2002; Stuijve, 2007). Moreover, the behavior of the RMSEA remains largely unravelled in the context of ordered-categorical data. Finally, as Hopwood and Donnellan (2010) have stressed, CFA could be a too stringent method to validate psychological instruments because of the complexity and incompleteness of most psychological theories, on the one hand, and method biases conveyed by psychological measures themselves, on the other hand. Therefore the factor structure of the baseline model was accepted. In their original work, Crocker, et al. (2003) also confirmed the configural equivalence across sex of their questionnaire, despite some minor indications of misspecification, but with the traditional method for continuous variables and ML estimator. Relying consequently on the relative fit indices for further invariance testing, the measurement model with equal loadings was assessed as fitting reasonably well both male and female data: CFI and TLI values were acceptable, and the relative chi-square difference statistic ($\Delta\chi^2/df$) was smaller than 2, while the RMSEA had not increased, indicating that the additional constraint on loadings did not significantly deteriorate the adequacy of the model. The seven latent factors (contingencies of self-worth) supposed to be measured by the questionnaire could thus be considered metrically equivalent for men and women in these groups. However, strong invariance was not altogether supported; among possible sources of variance, a sex bias in response was found partly responsible for some scalar differences.

As Byrne, Shavelson, and Muthén (1989) showed more than 20 years ago in a seminal didactic paper, there is more to the invariance issue than *complete* measurement invariance: *partial* measurement invariance may be sufficient for further latent mean structure testing. In the current study, the modification indices and residual covariances from the output of CFA configural invariance testing and the PCA showed that, while factors that would need cross-loadings or correlated error terms for their items seemed to be the same, it was not clear which particular items had to be selected for these adjustments. It would require a much more detailed and systematic study together with adequate replications before a stable basis for adjustments can be reached. Therefore, the authors chose to stay with the original 7-factor structure rather than making *ad hoc* adjustments.

We were able to confirm a reasonable fit of the original seven-factor structure of self-worth to the data gathered in male and female students with the newly translated CSWS, and also evidence for the metric invariance of this structure across sex was found. Further validation

work of the scale in French should nevertheless also take the following into account.

It is noteworthy that responses to items relating to the God's Love scale on self-worth indicated that the respondents posit themselves either in disagreement with any contingency on God's Love or as neutral. God's Love also had the lowest score range compared to other contingencies in Crocker's study (Crocker & Wolfe, 2001), but the current sample's mean score was far lower than theirs [2.49 ($SD=1.81$) compared to 4.2 ($SD=1.70$)]: this may relate to the particular student population or maybe to the Belgian French-speaking society as a whole. Moreover, this faith dimension of self-worth seemed to be quite unrelated to other contingencies in the sample. No further alternative measurement models, such as higher-order models, were considered in the present work, as the goal was basically confirmatory and the original scale was assumed to be sufficiently valid. However, it might be of methodological interest to assess how dropping the God's Love dimension would affect the fit of the truncated 6-factor model.

The data suggest that the partitioning of self-worth contingencies in seven different fields as originally proposed might deserve some reconsideration; of course, before making actual changes to the scale, these suggestions should first be explored further and the extent to which alternative measurement models could then be warranted should be a central issue of future work. More specifically, this study showed that responses to several items tapping supposedly separate domains of self-worth were related to an extent not accounted for by the specified model: this was mainly the case for items about Competition, Appearance, and Others' Approval contingencies. These "extra" (unexpected) links between items were due to shared variances attributable either to a specified non-primary factor (in case of significant modification indices for cross-loadings) or to an unknown (residual/error) covariance factor. For several of these covariances, given that the sample consisted of students, it makes sense that some items expressing competitive success were related to support from one's parents (Family Support) and that some Appearance items had competitive aspects. The pattern of estimated factor covariances indicated that most factors shared some variance, with the interesting exceptions of Virtue and mainly God's Love, which seemed rather independent of Appearance, Competition, and Others' Approval, leaving some room for testing a new measurement model in the future. It was shown however, that one such alternative model, positing that contingencies of self-worth can be organised into two factors contrasting internal and external sources of self-worth, was not adequate to fit the data. This confirmed Crocker, et al. (2003), who also demonstrated the seven-correlated-factor simple structure fitting data better than any substantively-based competing model.

Of course, method and respondent biases relating to the item wording and translation as well as cultural factors must also be taken into account in the validation of a questionnaire. Indeed, as van de Vijver and Leung (2000) and also Byrne and Watkins (2003) have stressed, unless the cross-cultural equivalence of the instrument, here the CSWS, has been ascertained (which was not part of the present work), factorial validation emanating from empirical work on the original American scale does not benefit the French version *de facto*. Actually, it cannot be excluded that some of the misfits noticed in the current CFA testing emanate from either problematic linguistic inconsistencies with the English version or particular attitudes of non-American French-speaking respondents or both. Note that both the current study and those of Crocker (2001, 2003) use comparable student samples, which rules out this kind of population contributions to model misfits.

Finally, the suggested sex differences in mean estimates of self-worth contingencies in the current study are surprising compared with results from other studies with the same questionnaire. Men seemed more likely to endorse Competition and Virtue, and less likely to endorse God's Love than were women. The importance of Competition for men was not surprising; neither was the apparent lesser importance of God's Love. Virtue contingency is also stronger for men than for women. This higher dependency on principles can be related to "Thinking" (vs. "Feeling"), a component from the Myers-Briggs Type Indicator (MBTI) typology (Briggs Myers, 1980). Thinking means following principles for decision making, whereas Feeling means that one's decisions are based on connections with other people. Gender differences seem to exist with respect to Thinking vs. Feeling, with 60% of men classified at the Thinking side and 60% of women classified at the Feeling side (Fox-Hines & Bowersock, 1995). These differences are not large and only significant at the .05 level, but they are a possible explanation for the findings: men in the sample may have understood the Virtue items in the Thinking sense. Given that strong measurement invariance was not established, but only metric invariance (invariance of what is being measured), the factor scores and thus their estimated means can be better trusted than the means of the summary scores as reported in other studies. A possible discrepancy between factor scores and summary scores as a way to test group differences in the seven contingency domains is therefore an interesting topic of further study.

Though the sample was large, the size was not determined *a priori* by a Monte Carlo study as now commonly recommended (Brown, 2006). Indeed, as stressed by Muthén and Muthén (2002) among others, the minimal sample size of a study depends on many factors, including the size and complexity of the model, the distribution and reliability of the variables, the amount of

missing data, and the strength of the relationships between variables. It is only very recently, to the authors' knowledge, that such approach was for the first time applied in the context of categorical variable methodology with a WLSMV estimator (Myers, Ahn, & Jin, 2011). This important aspect of model evaluation through CFA should certainly be taken into account beforehand in the design of future studies.

REFERENCES

- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003) Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, 4, 1-44.
- Beauducel, A., & Herzberg, P. Y. (2006) On the performance of maximum likelihood versus means and variance adjusted weighted least square estimation in confirmatory factor analysis. *Structural Equation Modeling*, 13, 186-203.
- Bentler, P. M. (1990) Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M. (2010) SEM with simplicity and accuracy. *Journal of Consumer Psychology*, 20, 215-220.
- Bollen, K. A. (2002) Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-34.
- Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton P., & Chen, F. (2007) Latent variable models under misspecification: two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods and Research*, 36, 48-86.
- Briggs Myers, I. (1980) *Gifts differing: understanding personality type*. Mountain View, CA: Davies-Black.
- Brown, T. A. (2006) *Confirmatory factor analysis*. New York: Guilford Press.
- Browne, M. W., & Cudeck, R. (1993) Alternate ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage. Pp. 136-162
- Browne, M. W., MacCallum, R. C., Kim, C-T., Andersen, B. L., & Glaser, R. (2002) When fit indices and residuals are incompatible. *Psychological Methods*, 7, 403-421.
- Byrne, B. (2008) Testing for multiple group equivalence of a measuring instrument: a walk through the process. *Psicothema*, 20, 872-882.
- Byrne, B. (2012) *Structural equation modeling with Mplus – basic concepts, applications, and programming*. New York: Taylor & Francis/Routledge.
- Byrne, B., Shavelson, R. J., & Muthén, B. (1989) Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466.
- Byrne, B., & Watkins, D. (2003) The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008) An empirical evaluation of the use of fixed cut-off points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, 36, 462-494.
- Crocker, J., & Luhtanen, R. K. (2003) Level of self-esteem and contingencies of self-worth: unique effects on academic, social, and financial problems in college freshmen. *Personality & Social Psychology Bulletin*, 29, 701-712.
- Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003) Contingencies of self-worth in college students: theory and measurement. *Journal of Personality & Social Psychology*, 85(5), 894-908.
- Crocker, J., & Wolfe, C. T. (2001) Contingencies of self-worth. *Psychological Review*, 108(3), 593-623.
- Fox-Hines, R., & Bowersock, R. B. (1995) ISFJ, ENTJ, MBTI; what's it all about. *Business and Economic Review*, 41(2), 3-8.
- Gentile, B., Grabe, S., Dolan-Pascoe, B., & Wells, B. E. (2009) Gender differences in domain-specific self-esteem: a meta-analysis. *Review of General Psychology*, 13, 34-45.
- Gregorich, S. E. (2006) Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44 (11, Suppl. 3), S78-S94.
- Hayduk, L. A., Cummings, G. C., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007) Testing! Testing! One, two, three—testing the theory in structural equation models. *Personality & Individual Differences*, 42, 841-850.
- Hopwood, C. J., & Donnellan, M. B. (2010) How should the internal structure of personality inventories be evaluated? *Personality & Social Psychology Review*, 14(3), 332-346.
- Hu, L. T., & Bentler, P. M. (1999) Cut-off criteria for fit indices in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jackson, D. L., Gillasp, J. A., Jr., & Purc-Stephenson, R. (2009) Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological Methods*, 14, 6-23.
- Josephs, R. A., Markus, H. R., & Tafarodi, R. W. (1992) Gender and self-esteem. *Journal of Personality and Social Psychology*, 63, 391-402.
- Kaniskan, B. (2011) *The performance of RMSEA in models with ordinal measures*. Paper presented at the 1st Annual Modern Modeling Methods (M³) Conference, Storrs, CT, May 25-26.
- Luhtanen, R. K., & Crocker, J. (2005) Alcohol use in college students: effects of level of self-esteem, narcissism, and contingencies of self-worth. *Psychology of Addictive Behaviors*, 19(1), 99-103.
- Lubke, G. H., & Muthén, B. O. (2004) Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, 11(4), 514-534.
- MacCallum, R. C. (2003) Working with imperfect models. *Multivariate Behavioral Research*, 38, 113-139.
- Maricutoiu, L. P., Macsinga, I., Rusu, S., Virga, D., & Sava, F. A. (2012) Adaptation and validation of the contingencies of self-worth scale on a Romanian student sample. *Cognition, Brain, Behavior*, 16, 121-138.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004) In search of golden rules: comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320-341.
- Meredith, W. (1993) Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Millsap, R. E., & Yun-Tein, J. (2004) Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479-515.
- Muthén, B., & Asparouhov, T. (2002) *Latent variable analysis with categorical outcomes: multiple-group and growth modeling in Mplus*. Webnote version 5, December 9. Retrieved November 16, 2011, from <http://www.statmodel.com/examples/webnotes.shtml>.
- Muthén, L. K., & Muthén, B. O. (1998-2004) *Mplus user's guide*. (3rd ed.) Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (1998-2009) *Mplus user's guide*. (5th ed.) Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002) How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599-620.

- Myers, N. D., Ahn, S., & Jin, Y. (2011) Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: a Monte Carlo approach. *Research Quarterly for Exercise & Sport*, 82, 412-423.
- Park, L. E., & Crocker, J. (2005) Interpersonal consequences of seeking self-esteem. *Personality & Social Psychology Bulletin*, 31, 1587-1598.
- Park, L. E., Crocker, J., & Kiefer, A. K. (2007) Contingencies of self-worth, academic failure, and goal pursuit. *Personality & Social Psychology Bulletin*, 33, 1503-1517.
- Prudon, P. (2011) Goodness-of-fit indices in confirmatory factor analysis: an unreliable guide. Unpublished manuscript. Retrieved December 18, 2012, from <http://www.fzp-cursussen.nl/GOF.pdf>, SPSS versions 17.0-19.0 (2008-2010), SPSS Inc., Chicago, IL.
- Stefanone, M. A., Lackaff, D., & Rosen, D. (2011) Contingencies of self-worth and social-networking-site behavior. *Cyberpsychology, Behavior and Social Networking*, 14, 41-49.
- Stuive, I. (2007) A comparison of confirmatory factor analysis methods: oblique multiple group method versus confirmatory common factor method. Unpublished Ph.D. thesis, Univer. of Groningen, the Netherlands. Retrieved October 12, 2012, from <http://dissertations.ub.rug.nl/faculties/gmw/2007/i.stuive>.
- Tomarken, A. J., & Waller, N. G. (2005) Structural equation modeling: strengths, limitations and misconceptions. *Annual Review of Clinical Psychology*, 1, 31-65.
- Tucker, L. R., & Lewis, C. (1973) A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- van der Sluis, S., Vinkhuyzen, A. A. E., Boomsma, D. I., & Posthuma, D. (2010) Differences in adults' motivation to achieve. *Intelligence*, 38, 433-446.
- van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, 31, 33-51.